NMR spectra database for on-flight identification of HPLC-SPE-NMR data

Johansen, Kenneth

Publication date:
2011

Document version
Early version, also known as pre-print

FACULTY OF PHARMACEUTICAL SCIENCES
UNIVERSITY OF COPENHAGEN

NMR spectra database for on-flight identification of HPLC-SPE-NMR data

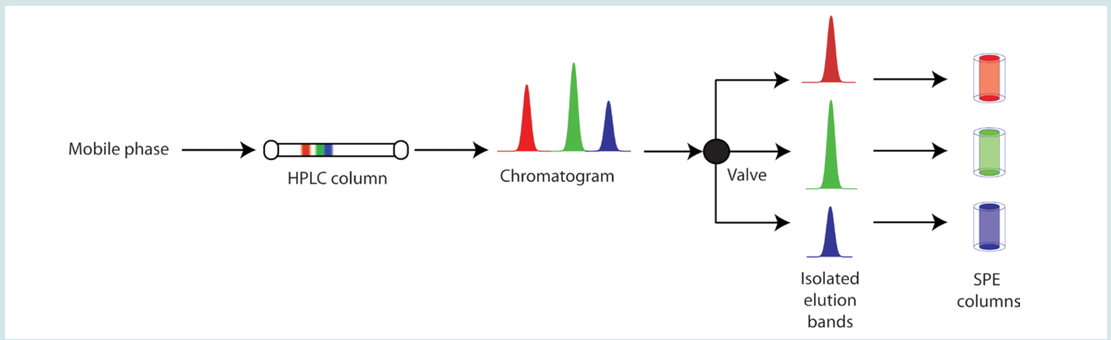
Kenneth Johansen[1*], Nils T. Nyberg[1], Manfred Spraul[2], Jerzy W. Jaroszewski[1]
[1]Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen,Universitetsparken 2, DK-2100
[2]NMR applications Group, Bruker Biospin GmbH, Silberstreifen 4, DE-76287
*E-mail: kj@farma.ku.dk

Introduction

Current developments within sensitivity of NMR spectroscopy combined with hyphenated methods such as HPLC-SPE-NMR have eased the work of natural product chemists. The fact that we are now capable of automatically acquiring NMR data of even minor metabolites in the ng-range [1] makes dereplication based on structure elucidation viable. This work presents such a dereplication tool based on the information obtained with a rapidly acquired ^{1}H NMR spectrum.

HPLC-MS-SPE-NMR

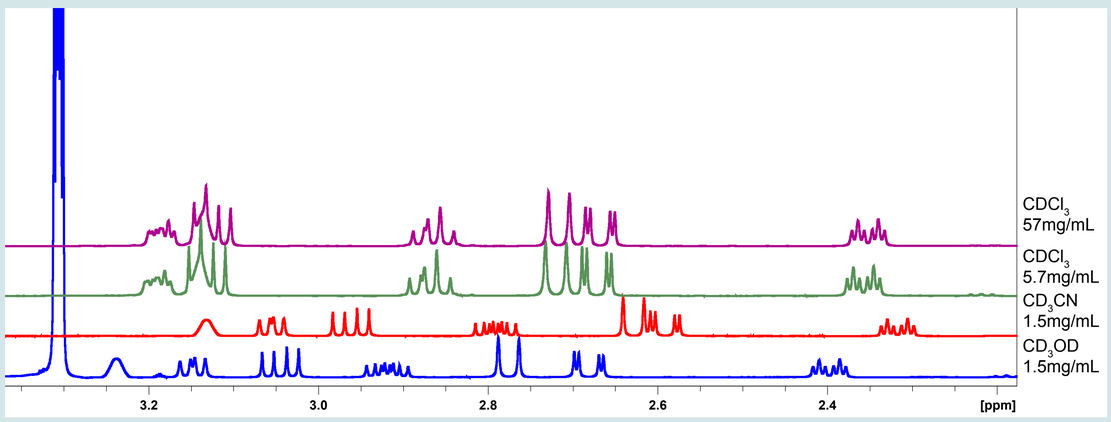
Advances within the area of NMR hyphenation have led to implementation of an on-line solid phase extraction (SPE) step between the outlet of HPLC and the NMR. By lowering the eluotropic strength of the eluent through addition of water, trapping of metabolites onto SPE cartridges can be triggered by, e.g,. ion count (mass detection) or absorption (UV detection). After drying, the metabolites can be transferred with deuterated solvents to either a flowprobe or into NMR tubes in automation.



One of the numerous advantages of the HPLC-SPE-NMR setup over other hyphenated NMR techniques [2][3][4] is the increased sensitivity obtainable by doing repeated trappings. Our lab has recently been equipped with a cryogenically cooled 1.7 mm probe installed in our 600 MHz magnet, making it possible to acquire NMR data of a vast majority of even minor metabolites in a crude extract in automation. Although the hyphenated system is equipped with a high resolution mass spectrometer - traditionally used for dereplication - we propose the use of the easily acquired, structurally information-rich ^{1}H spectra instead.

Challenges with a ^{1}H dereplication tool

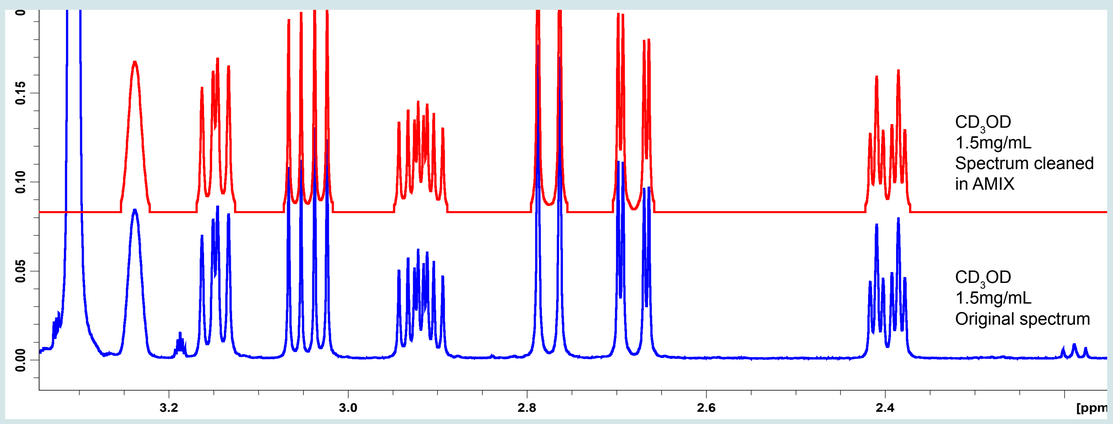
The strength of NMR as a dereplication tool is that a signal not only tells something about a given proton, but also about its chemical environment. This strength is unfortunately also a weakness since many factors such as solvent, temperature, ionic strength, shimming, and concentration can cause changes in chemical shift values or signal appearances as illustrated below with strychnine.



Another challenge with using NMR data for dereplication purposes compared to the numerical value used for MS-dereplication is the need for reference spectra. With the data storage capacity available today, the database size is not an issue, but to ensure time-efficient database seaches, noise and artefacts need to be removed for the matching program to only work on regular signals.

Spectral preparation

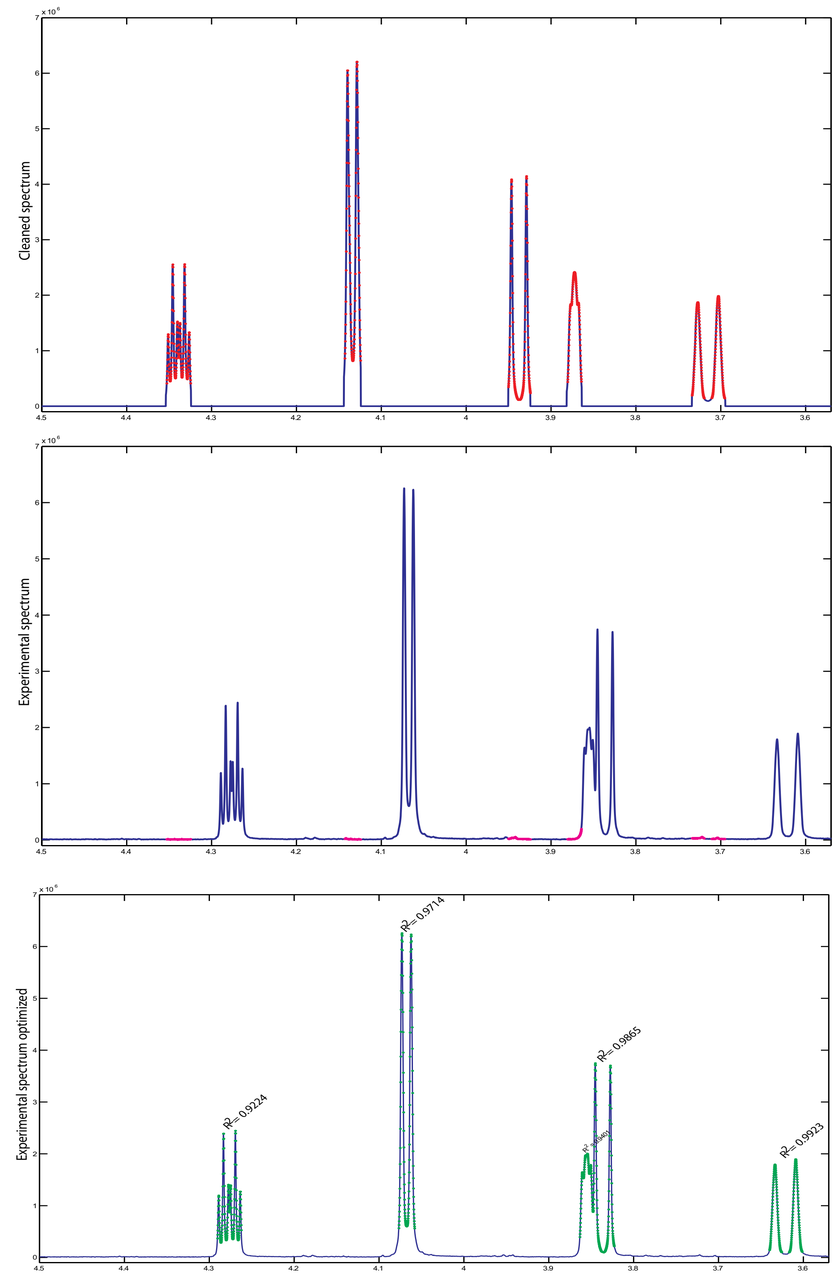
The spectral preparation is conducted using the AMIX platform by Bruker Biospin. This software is primarily developed as a metabolomics tool with a built-in database feature, in which 1D and 2D spectra can be manually cleaned for noise, artifacts and solvent peaks by exchanging these with zeros, as shown below.



Spectral matching

The dereplication tool was developed in the MATLAB environment, which offers great flexibility. The tool is still under development and therefor currently lacks a graphical user interface. It divides the chosen reference spectra into subspectra, calculates the best match for each signal and returns an overall match factor as well as a graphical presentation of each matched signal for visual inspection.
The dereplication tool currently accepts user inputs regarding the maximum allowed shift of resonances, maximum amount of data points used for matching, and which databases to match against.

Below is given a graphical presentation showing how the subspectra matching is functioning in even an extreme case, where the analyte have been dissolved in CD_3OD and CD_3CN.



The points used for the match are indicated with red dots in the cleaned spectrum. The magenta dots indicate the match that would be achieved by simply superimposing the cleaned spectrum onto the experimental data, while the green dots represent the same points after the subspectra match routine has been run.

Results and perspective

In this study, we show how ^{1}H NMR data hold a great potential for automatically dereplicating crude extracts of natural products. The dereplication tool we have developed in the MATLAB environment is based on a subspectra match routine that takes resonance shifts, often encountered in NMR spectroscopy, into account.

With the amount of structural information obtainable from a ^{1}H spectrum, we believe that, with a growing database size, this tool will be able to not only dereplicate but also be a valuable help as a structure elucidation tool or for classification of compounds into compound classes.

References

[1] Hilton BD, Martin GE, *J. Nat. Prod.* 2010; 73: 1465-69
[2] Kesting JR, Johansen KT, Jaroszewski JW *in* Biomolecular NMR Spectroscopy, Chapter 19, IOS Press, 2011; 3
[3] Jaroszewski JW, *Planta Med* 2005; 71: 795-82
[4] Exarchou V *et al. Current Trends in Analysis and Characterisation* 2006: 143-155

Acknowledgements

The authors would like to thank Bruker Biospin GmbH and the Drug Research Academy for co-financing the PhD fellowship to Kenneth Johansen. Dr. Christian Fischer and Dr. Klaus-Peter Neidig are thanked for valuable help in building the database.